Cell PRESS
Structure
Article

Toward a Structure Determination Method for Biomineral-Associated Protein Using Combined Solid-State NMR and Computational Structure Prediction

David L. Masica,[1] Jason T. Ash,[3] Moise Ndao,[3] Gary P. Drobny,[3,*] and Jeffrey J. Gray[1,2,*]
[1]Program in Molecular Biophysics
[2]Department of Chemical and Biomolecular Engineering
Johns Hopkins University, Baltimore, MD 21218, USA
[3]Department of Chemistry, University of Washington, Seattle, WA 98195, USA
*Correspondence: drobny@chem.washington.edu (G.P.D.), jgray@jhu.edu (J.J.G.)
DOI 10.1016/j.str.2010.09.013

SUMMARY

Protein-biomineral interactions are paramount to materials production in biology, including the mineral phase of hard tissue. Unfortunately, the structure of biomineral-associated proteins cannot be determined by X-ray crystallography or solution nuclear magnetic resonance (NMR). Here we report a method for determining the structure of biomineral-associated proteins. The method combines solid-state NMR (ssNMR) and ssNMR-biased computational structure prediction. In addition, the algorithm is able to identify lattice geometries most compatible with ssNMR constraints, representing a quantitative, novel method for investigating crystal-face binding specificity. We use this method to determine most of the structure of human salivary statherin interacting with the mineral phase of tooth enamel. Computation and experiment converge on an ensemble of related structures and identify preferential binding at three crystal surfaces. The work represents a significant advance toward determining structure of biomineral-adsorbed protein using experimentally biased structure prediction. This method is generally applicable to proteins that can be chemically synthesized.

INTRODUCTION

Biomineralization is a process of crystal nucleation and growth controlled by bioorganic molecules such as proteins (Dove et al., 2003; Mann, 2001; Sigel et al., 2008). During crystal growth, proteins can locate and function at the liquid-solid surface phase boundary and accelerate (Elhadj et al., 2006), inhibit (Boskey et al., 1993; Shiraga et al., 1992), or shape crystal growth (Kensuke and Yoshiki, 2001; Sollner et al., 2003). Many organisms utilize biomineralization to fabricate the solid inorganic components of biogenic materials. These include a variety of structural and functional materials and the mineral phase of hard tissues. Some common biogenic materials include calcified minerals, magnetite, and biosilica. Deleterious biomineralization can result in pathologies such as kidney stones (Dussol et al., 1995; Ryall, 1996), dental calculus, and atherosclerosis (Dorozhkin and Epple, 2002).

The structure of biomineral-associated proteins cannot be determined by X-ray crystallography or solution nuclear magnetic resonance (NMR); as a result, high-resolution aspects of protein biomineralization are not well understood. Although many experimental methods exist for studying proteins adsorbed to solid surfaces, most can only resolve macroscopic features (Gray, 2004). Solid-state NMR (ssNMR) is uniquely suited for determining the distance between pairs of isotopically labeled atoms at the protein-surface interface (Goobes et al., 2007). High-resolution ssNMR data exists for at least two protein-biomineral complexes: statherin (Gibson et al., 2005; Goobes et al., 2006a; Stayton et al., 2003) and the leucine-rich amelogenin protein (Shaw et al., 2008). In solution-state NMR, 10–15 measurements are typically acquired at each residue during protein structure determination (Bowers et al., 2000); this is currently not tractable by ssNMR methods for surface-bound proteins. Although analogous multidimensional ssNMR studies of microcrystalline proteins can in principle obtain a comparable number of measurements per residue, biomaterial surface heterogeneities limit spectroscopic resolution and make obtaining a similar density of structural constraints less practical for surface-adsorbed proteins.

Recently, combined solution NMR-computational structure prediction methods have drastically reduced the amount of NMR data necessary to determine high-resolution protein structures in solution (Cavalli et al., 2007; Gong et al., 2007; Shen et al., 2008). In particular, the Rosetta structure prediction method was combined with chemical shift (Shen et al., 2008), nuclear Overhauser effect (Bowers et al., 2000), or residual dipolar coupling (Meiler and Baker, 2003; Rohl and Baker, 2002) NMR data. In all three cases, sets of protein structures predicted by Rosetta were at or near atomic-level accuracy. The use of high-resolution experimental data biases sampling to relevant conformation space and helps account for inaccuracies in computational energy functions. An approach combining high-resolution experiment and structure prediction has the potential to elucidate biomineral-associated protein structure to unprecedented resolution.

Previously, we developed Rosetta to predict the fold and orientation of a protein on a biomineral surface (RosettaSurface) (Masica and Gray, 2009). In addition, we developed and tested

Table 1. Solid-State NMR Measurements Used in First Round of Biased Structure Prediction

Measurement	Label	Distance (Å) or Angle (°)	Reference
pS_3 ϕ	$^{13}C'_2$-$^{13}C'_3$	−60° ± 10	Long et al., 2001
pS_3-F_7	$^{13}C'$-^{15}N	4.3 ± 0.2 Å	Long et al., 2001
K_6-HAp	$^{15}N\zeta$-^{31}p	5.6 ± 0.5 Å	Gibson et al., 2005
F_7-HAp	$^{13}C6$-^{31}p	6.9 ± 1.0 Å	Gibson et al., 2006
L_8 ϕ	$^{13}C'_8$-$^{13}C'_9$	−60° ± 9	Long et al., 2001
L_8-G_{12}	$^{13}C'$-^{15}N	−4.8 ± 0.4 Å	Long et al., 2001
G_{12} ϕ	$^{13}C'_{12}$-$^{13}C'_{13}$	−73° ± 3.6	Long et al., 2001
P_{23}-P_{33}	$^{13}C^{\beta}$-$^{13}C'$	8.8 ± 0.8 Å or 10.5 ± 1.0 Å	Goobes et al., 2006a
P_{23}-Y_{34}	$^{13}C^{\beta}$-$^{13}C'$	8.8 ± 0.8 Å or 10.5 ± 1.0 Å	Goobes et al., 2006a
P_{33}-Y_{34}	$^{13}C'$-$^{13}C'$	3.12 ± 0.13 Å	Goobes et al., 2006a
P_{33}-Y_{38}	$^{13}C'$-^{15}N	5.3 ± 0.5 Å	Goobes et al., 2006a
Y_{34} ϕ	$^{13}C'_{33}$-$^{13}C'_{34}$	−75° ± 15	Goobes et al., 2006a
Y_{34} ψ	$^{13}C'_{33}$-$^{13}C'_{34}$	−40° ± 10	Goobes et al., 2006a
Y_{34}-Y_{38}	$^{13}C'$-^{15}N	4.0 ± 0.5 Å	Goobes et al., 2006a

The residue for which the measurement was acquired, the placement of the isotopic label, and the determined distance (Å) or angle (°). Angle measurements were determined using $^{13}C'$ of that residue and $^{13}C'$ of the preceding residue. Multiple values for long-range C-terminal measurements indicate ambiguity in the experimental measurement.

structure prediction biased by ssNMR measurements. Here, we present an improved version of ssNMR-biased RosettaSurface structure prediction (RosettaSurface.NMR). RosettaSurface.NMR folds a protein on a biomineral surface using ssNMR data to bias structure prediction and has improved sampling of protein-surface intermolecular constraints.

As an example of the application of RosettaSurface.NMR, we attempt to determine the atomic-level structure of a 43-residue protein (salivary statherin) bound to its biologically relevant crystal surface, hydroxyapatite (HAp). We perform two iterations (rounds) of biased predictions. In round 1, we use previously published ssNMR measurements to predict the structure of statherin bound to a single HAp crystal surface. Structures from round 1 predictions are compared with six additional, recently published ssNMR measurements (Ndao et al., 2009; Ndao et al., 2010), and two new measurements. Results from the new and all previously published measurements bias a second round of RosettaSurface.NMR predictions for statherin adsorbed to five HAp crystal surfaces. Results from round 2 biased structure prediction suggest preferential adsorption at three of these HAp crystal surfaces. Finally, we perform more extensive sampling at one HAp crystal surface to produce a final set of proposed structures.

RESULTS

Statherin is low-molecular weight, highly charged, and proline-rich; these factors can contribute to flexibility in protein conformation. Similarly, surfaces can catalyze structural transitions in proteins. Therefore, we consider both the ensemble properties and individual structures in our analysis.

We performed two rounds of ssNMR-biased structure prediction, beginning first with published ssNMR measurements (Table 1). Because the HAp {001} crystal face is a primary growth plane (Simmer and Fincham, 1995) and statherin is a HAp growth inhibitor (Schwartz et al., 1992), the HAp {001} crystal face is a potential surface for statherin adsorption. To simplify calculations in the first round, we predicted statherin's structure only at the HAp {001} surface.

Round 1

Figure 1 shows protein intramolecular (Figure 1A) and protein-surface intermolecular (Figure 1B) contacts for the 100 structures with the lowest constraint energy (i.e., most in agreement with experimental measurements) from round 1 predictions. Predicted contacts are useful for suggesting measurements for subsequent rounds of ssNMR and for indicating the predicted structure of the ensemble. During round 1 predictions, the constraint weight *w* (see Experimental Procedures) was set to 10 kcal/mol.

In Figure 1A, the dense regions of *i* to *i* + 4 contacts in the N-terminal domain suggest helical secondary structure from residues 3 to 14, with some helical fraying from residues 11 to 14. The *i* to *i* + 4 contacts between residues 31 and 39 also suggest helical structure. There are few contacts made between residues that are greater than five residues apart in primary sequence, suggesting little tertiary structure. The contacts made in the segment intervening residues 23 and 34 arise from the ssNMR measurements used to bias structure prediction (see Table 1).

In Figure 1B, high-frequency residue-surface contacts exist only in statherin's N-terminal domain, indicating that HAp binds primarily to that domain. Asp1, Sep3, and to a lesser extent Sep2 are all located at the interface (Sep is an abbreviation for phosphorylated serine). Isotopic labeling of phosphoserine residues is avoided because it is difficult to distinguish between ssNMR signal from HAp surface phosphates and Sep side-chain phosphates. Glu5 is predicted to adsorb frequently whereas Glu4 is predicted to point away from the surface. The Glu5 side-chain carboxyl oxygens are predicted to be ~2.5 Å from a HAp calcium atom; Glu4 is predicted to be ~8 Å from the HAp surface. Arg9, Arg10, and Arg13 are also predicted to bind HAp. The C-terminal domain does not bind, even though there is a negatively charged glutamic acid residue in that region (Glu26). This may be surprising considering the high affinity of acidic amino acids for HAp.

Round 2

We chemically synthesized eight new constructs labeled at unique positions to compare with round 1 predictions. NMR measurements for six of these labeled constructs were recently reported in the literature and two are reported here for the first time. The measurements include intermolecular distances from HAp phosphate phosphorus atoms to residues Glu4, Glu5, Glu26 (Ndao et al., 2009), Arg9, Arg10, Arg13 (Ndao et al., 2010), Phe14, and Pro28 (new). The orientation of these residues at the interface should help reveal the bound orientation of the statherin-HAp complex, and will address round 1 predictions including the disproportionate binding of neighboring acidic residues, binding of basic residues, and lack of binding in the

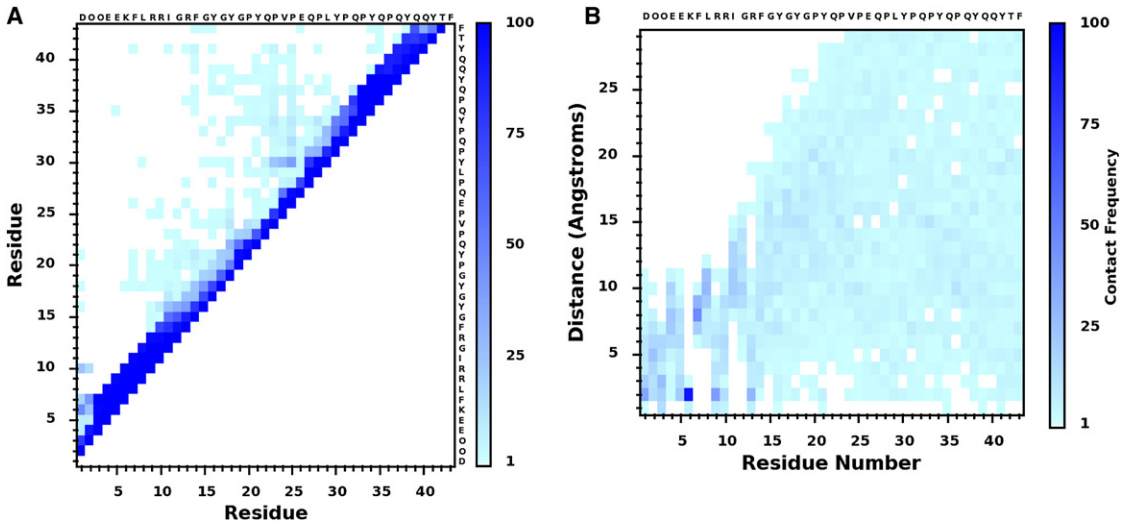

Figure 1. Ensemble Structure of Statherin Adsorbed to HAp {001} Surface Determined from the First Round of Biased Structure Prediction
(A) Pairwise statherin intramolecular residue-residue contacts and (B) pairwise residue-surface distances for the 100 structures with smallest $E_{constraint}$ (see Experimental Procedures). An intramolecular residue-residue contact is declared if two residues have an inter-residue atomic pair within 4 Å. A residue-surface distance reflects the closet atomic contact for that residue and the closest surface atom. Note for comparing contacts with ssNMR measurements: the intra- or intermolecular atomic pairs that constitute a contact are not necessarily the same atoms that were isotopically labeled for ssNMR measurements.

C-terminal domain. Also, the high number of protein-surface intermolecular constraints should allow us to address the possibility of crystal face specificity by performing biased simulations at multiple HAp crystal faces.

Table 2 shows results for the new and recently published set of ssNMR distance measurements. These measurements indicate that Glu5, Arg9, and Arg10 are located at the HAp surface (53%, 72%, and 44% of round 1 low-energy ensemble within 6 Å of surface, at these residues, respectively) and that Glu4, Phe14, Glu26, and Pro28 do not bind HAp (10%, 20%, 9%, and 12% of round 1 low-energy ensemble within 6 Å of surface, at these residues, respectively), in agreement with round 1 predictions. This agreement suggests convergence between ssNMR and RosettaSurface.NMR, at least in the ssNMR measured regions. However, Round 1 calculations predicted strong HAp binding via Arg13, whereas ssNMR measurements show a HAp-Arg13 distance of >7 Å.

Round 2 predictions were biased using all ssNMR measurements from Tables 1 and 2. We performed round 2 ssNMR-biased structure prediction at five HAp crystal faces: {001}, {010}, two differentially terminated {100} faces ({100}-T1 and {100}-T2), and {101} (see Figure 2). The {100}-T2 surface terminates with a higher density of calcium ions than the {100}-T1 surface (Figures 2D and 2E respectively). All other HAp surfaces were cut to expose approximately neutral (mixed charged) surfaces (Figure 2). These faces were chosen because they are expressed in stable HAp crystals and have a range of surface geometries (Astala and Stott, 2008).

Table 2. Solid-State NMR Measurements Used in Second Round of Biased Structure Prediction

Measurement	Label	ssNMR Distance (Å)	Reference
E_4-HAp	$^{13}C^{\delta}$-^{31}p	>7.28	Ndao et al., 2009
E_5-HAp	$^{13}C^{\delta}$-^{31}p	4.2 ± 0.3 Å	Ndao et al., 2009
R_9-HAp	$^{13}C^{\zeta}$-^{31}p	4.5 ± 0.3 Å	Ndao et al., 2010
R_{10}-HAp	$^{13}C^{\zeta}$-^{31}p	4.4 ± 0.2 Å	Ndao et al., 2010
R_{13}-HAp	$^{13}C^{\zeta}$-^{31}p	>7.28	Ndao et al., 2010
F_{14}-HAp	$^{13}C6$-^{31}p	>7.28	New
E_{26}-HAp	$^{13}C^{\delta}$-^{31}p	>7.28	Ndao et al., 2009
P_{28}-HAp	$^{13}C'$-^{31}p	>7.28	New

The residue for which the measurement was acquired, the placement of the isotopic label, and the determined distance (Å).

At each surface, the predictions were divided into 11 runs of ~9100 candidate structures. For each of the 11 runs at each surface, the constraint weight w was set to a number between 0 and 1 kcal/mol evenly divisible by 0.1 (i.e., 0, 0.1, 0.2 …. 1). We used this approach to test whether the set of constraints is easier to match at a particular statherin-HAp interface. This result would manifest as a prediction at one or more HAp surfaces exhibiting smaller $E_{constraint}$ versus weight relative to predictions at the other HAp surfaces. This approach has the potential to probe surface specificity at protein-surface interfaces.

Figure 3 shows $E_{constraint}$ versus weight for structures resulting from round 2 ssNMR-biased predictions. As expected, $E_{constraint}$ decreases with increasing weight at all five HAp surfaces. For each weight, $E_{constraint}$ is smallest at the {001}, {010}, and {100}-T1 surfaces. That is, it is easiest for RosettaSurface to create structures matching the experimental constraints when statherin binds the {001}, {010}, and {100}-T1 surfaces, suggesting that these surfaces are more likely bound in the ssNMR experiments. Convergence between prediction and experiment was slowest at the {100}-T2 and {101} surface; the {100}-T2 surface has a high calcium concentration (Figure 2E) and the {101} surface has a unique geometry (Figure 2F).

Figure 2 shows dimensions for the smallest periodic motif of open phosphate clusters at each HAp surface used for these predictions (phosphate clusters on the {100-T2} surface are

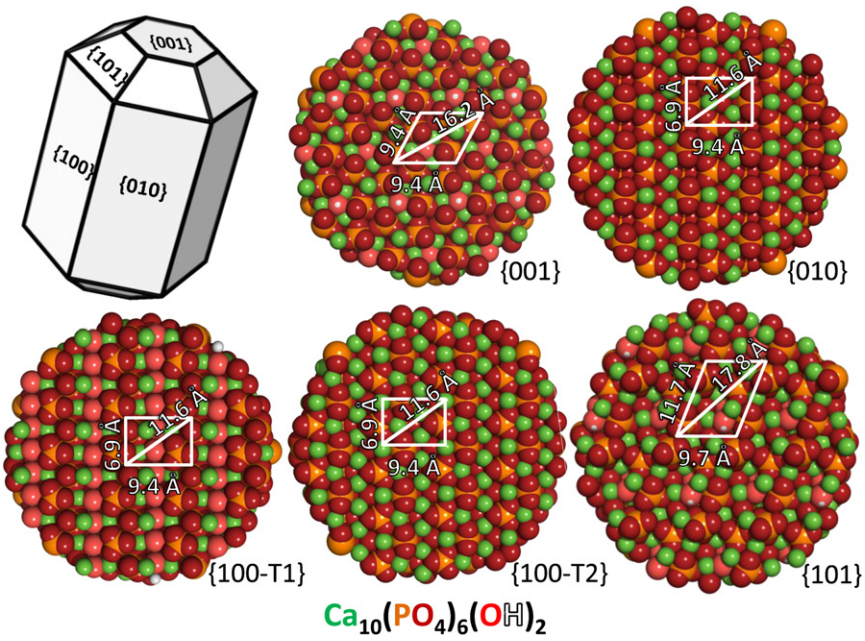

Figure 2. Hydroxyapatite Model

(A) Schematic of hexagonal HAp showing the five crystal faces (two differentially terminated surfaces at the {100} crystal plane) and the smallest periodic motif of open phosphate clusters at the (B) {001}, (C) {010}, (D) {100}-T1, (E) {001}-T2, and (F) {101} used for biased RosettaSurface.NMR structure prediction.

occupied by calcium atoms). The complementarity of these motifs to the motif of statherin's binding domain (and hence the relative positions of binding residues) plays an important role where convergence between ssNMR and structure prediction are concerned (Figure 3). Motifs at {001} and {010} surfaces best facilitates binding of Glu5, Lys6, Arg9, and Arg10 (Table 3). The periodic motif of phosphates at the {100}-T1 surface is similar to that of the {010} surface ({010} and {100} crystal lattices are identical aside from the direction of hydroxyl groups); however, the terminations chosen here have different calcium positions. At the {100}-T1 surface, binding via statherin basic residues is preferred at the expense of binding via Glu5 (Table 3), suggesting preferential binding at the {001} and {010} surfaces. The {101} surface has a unique geometry, whose dimensions are larger than other surfaces used for these predictions. Binding via basic residues and Glu5 are compromised at the {101} surface. The {100}-T2 surface is identical to the {010} surface with the addition of a calcium atom deposited into each phosphate cluster. Binding via Glu5 is among the best at that {100}-T2 surface, but binding of basic residues is disproportionately low at this surface. For statherin residues not directly involved in binding HAp, agreement between structure prediction and ssNMR is similar at each face.

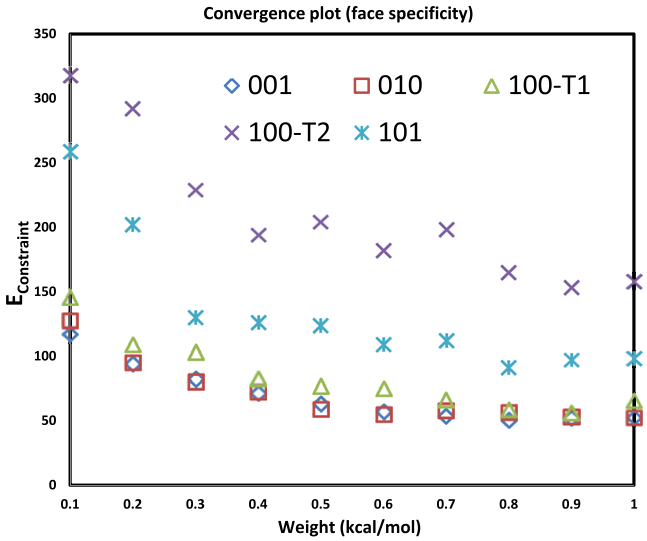

Figure 3. $E_{constraint}$ Versus Weight at Five Different HAp Surfaces

For the last phase of biased RosettaSurface.NMR structure determination, we generated 10^5 structures of HAp {001}-bound statherin using all ssNMR measurements (Table 3) and a weight of 10 kcal/mol. Figure 4 shows ensemble structural statistics for the 100 structures with smallest $E_{constraint}$ from this final phase of structure determination. Figure 4A shows a stable helix from residue 4–11 and less populated helix from residue 34–39. The ensemble shows turn and extended structure for statherin's middle segment. In Figure 4B, the residue-surface contact map shows fewer, more populated bins compared with predictions from round 1 (Figure 1B). This increased resolution shows the benefit incurred from biasing RosettaSurface.NMR predictions with the additional ssNMR measurements in Table 2. In particular, Glu5, Lys6, Arg9, and Arg10 are located at the HAp surface, and Glu4, Phe7, Leu8, Ile11, and Gly12 have defined positions ~8 Å from the surface.

Figure 5 shows a representative structural model from the 100 structures with smallest $E_{constraint}$ from the final phase of round 2 structure determination. This structure was chosen because it represents the dominant ensemble conformation (Figure 4) and closely matches ssNMR measurements (Table 3). Figure 5A shows a global view of the statherin-HAp complex. For that model, Figures 5B–5D show predicted distances and angles for which biasing was applied. These figures directly show the extent of biasing at specific regions of the complex and the structure that was predicted in those regions. Table 3 shows all experimental measurements from rounds 1 and 2 and the corresponding predictions from the representative model (Figure 5).

To assess the level of dispersion between the models, Figure 6 shows the local root-mean-square deviation (RMSD) in a representative structural model (Figure 5) relative the 100 structures with smallest $E_{constraint}$. This figure is complementary to the

Table 3. Comparison of All ssNMR Measurements, Ensemble Statistics at Each HAp Surface from Round 2 Biased Structure Prediction, and Predicted Angles and Distances from a Representative Structure from the Final Biased Structure Prediction at the HAp {001} Surface

Measurement	Label	Distance (Å) or Angle (°)	{001}	{010}	{100}−T1	{100}−T2	{101}	Rep Model
pS_3 ϕ	$^{13}C'_2$-$^{13}C'_3$	−60 ± 10°	−63.9 ± 12.0° (67)	−62.5 ± 14.8° (64)	−62.9 ± 9.8 °(73)	−63.3 ± 13.0° (68)	−63.1 ± 14.0° (66)	−64.8°
pS_3-F_7	$^{13}C'$-^{15}N	4.3 ± 0.2 Å	4.3 ± 0.3 Å (61)	4.3 ± 0.3 Å (58)	4.3 ± 0.4 Å (54)	4.2 ± 0.4 Å (53)	4.3 ± 0.3 Å (54)	4.5 Å
E_4-HAp	$^{13}C^{\delta}$-^{31}p	>7.28 Å	11.7 ± 1.4 Å (98)	11.7 ± 1.3 Å (99)	11.8 ± 1.3 Å (100)	12.3 ± 1.7 Å (99)	12.1 ± 1.6 Å (97)	13.1 Å
E_5-HAp	$^{13}C^{\delta}$-^{31}p	4.25 ± 0.27 Å	5.0 ± 0.5 Å (18)	5.0 ± 0.5 Å (22)	5.5 ± 0.4 Å (3)	5.4 ± 0.8 Å (22)	5.5 ± 0.7 Å (7)	4.5 Å
K_6-HAp	$^{15}N^{\varepsilon}$-^{31}p	4.25 ± 0.56 Å	4.3 ± 0.8 Å (64)	4.4 ± 0.9 Å (63)	4.1 ± 0.8 Å (66)	5.3 ± 1.3 Å (32)	4.9 ± 1.2 Å (45)	4.6 Å
F_7-HAp	^{13}C6-^{31}p	>6.75 Å	11.3 ± 1.4 Å(100)	11.1 ± 1.6 Å (100)	11.3 ± 1.5 Å (100)	10.9 ± 1.8 Å (98)	10.9 ± 1.7 Å (97)	9.5 Å
L_8 ϕ	$^{13}C'_8$-$^{13}C'_9$	−60 ± 9°	−62.7 ± 6.2° (84)	−62.9 ± 5.9° (83)	−61.9 ± 6.1° (83)	−62.3 ± 8.1° (74)	−63.5 ± 6.4° (80)	−74.4°
L_8-G_{12}	$^{13}C'$-^{15}N	4.8 ± 0.4 Å	4.3 ± 0.5 Å (26)	4.2 ± 0.4 Å (25)	4.2 ± 0.3 Å (22)	4.3 ± 0.8 Å (28)	4.2 ± 0.3 Å (24)	5.0 Å
R_9-HAp	$^{13}C^{\zeta}$-^{31}p	4.62 ± 0.29 Å	4.6 ± 0.4 Å (69)	4.7 ± 0.3 Å (72)	4.6 ± 0.4 Å (62)	5.5 ± 1.0 Å (29)	4.8 ± 0.5 Å (63)	4.4 Å
R_{10}-HAp	$^{13}C^{\zeta}$-^{31}p	4.53 ± 0.16 Å	4.8 ± 0.3 Å (30)	4.6 ± 0.4 Å (35)	4.8 ± 0.3 Å (33)	5.1 ± 0.4 Å (17)	4.9 ± 0.3 Å (20)	4.5 Å
G_{12} ϕ	$^{13}C'_{12}$-$^{13}C'_{13}$	−73 ± 3.6°	−71.4 ± 8.2° (26)	−71.4 ± 9.2° (38)	−71.4 ± 7.4° (29)	−72.2 ± 8.3° (36)	−73.5 ± 9.7° (34)	−72.5°
R_{13}-HAp	$^{13}C^{\zeta}$-^{31}p	>7.28 Å	7.6 ± 2.5 Å (44)	7.1 ± 1.7 Å (38)	7.4 ± 2.2 Å (40)	7.0 ± 2.5 Å (28)	6.9 ± 1.5 Å (35)	17.0 Å
F_{14}-HAp	^{13}C6-^{31}p	>7.28 Å	12.4 ± 2.8 Å (97)	12.3 ± 2.4 Å (99)	12.4 ± 2.6 Å (96)	10.6 ± 2.7 Å (92)	10.6 ± 2.6 Å (91)	9.5 Å
P_{23}-P_{33}	$^{13}C^{\beta}$-$^{13}C'$	8.8 ± 0.8 Å or 10.5 ± 1.0 Å	12.8 ± 2.5 Å (26)	12.6 ± 3.1 Å (30)	12.9 ± 2.5 Å (24)	12.8 ± 3.2 Å (25)	13.5 ± 2.8 Å (26)	12.4 Å
P_{23}-Y_{34}	$^{13}C^{\beta}$-$^{13}C'$	8.8 ± 0.8 Å or 10.5 ± 1.0 Å	13.0 ± 2.6 Å (26)	13.0 ± 3.2 Å (26)	13.3 ± 2.5 Å (25)	13.3 ± 3.1 Å (27)	13.6 ± 2.9 Å (20)	12.0 Å
E_{26}-HAp	$^{13}C^{\delta}$-^{31}p	>7.28 Å	21.3 ± 8.5 Å (96)	21.8 ± 8.4 Å (98)	19.9 ± 9.7 Å (89)	20.9 ± 8.0 Å (96)	20.9 ± 8.0 Å (94)	27.9 Å
P_{28}-HAp	$^{13}C^{\delta}$-^{31}p	>7.28 Å	21.7 ± 8.7 Å (98)	23.2 ± 8.5 Å (99)	22.2 ± 8.4 Å (99)	21.7 ± 7.6 Å (98)	21.6 ± 7.4 Å (96)	34.3 Å
P_{33}-Y_{34}	$^{13}C'$-$^{13}C'$	3.12 ± 0.13 Å	3.1 ± 0.1 Å (94)	3.0 ± 0.1 Å (95)	3.1 ± 0.1 Å (92)	3.1 ± 0.1 Å (90)	3.1 ± 0.1 Å (97)	3.0 Å
P_{33}-Y_{38}	$^{13}C'$-^{15}N	5.3 ± 0.5 Å	6.4 ± 0.4 Å (7)	6.3 ± 0.5 Å (14)	6.4 ± 0.5 Å (10)	6.2 ± 0.7 Å (20)	6.4 ± 0.8 Å (13)	6.0 Å
Y_{34} ϕ	$^{13}C'_{33}$-$^{13}C'_{34}$	−75 ± 15°	−67.2 ± 8.5° (80)	−64.7 ± 7.2° (76)	−65.9 ± 8.3° (77)	−65.8 ± 7.0° (78)	−66.7 ± 8.0° (80)	−65.0°
Y_{34} ψ	$^{13}C'_{33}$-$^{13}C'_{34}$	−40 ± 10°	−36.8 ± 8.2° (83)	−38.1 ± 7.5° (84)	−37.7 ± 8.3° (75)	−38.2 ± 8.7° (79)	−39.1 ± 9.1° (73)	−45.1°
Y_{34}-Y_{38}	$^{13}C'$-^{15}N	4.0 ± 0.5 Å	4.3 ± 0.6 Å (78)	4.2 ± 0.6 Å (84)	4.2 ± 0.5 Å (85)	4.2 ± 0.5 Å (84)	4.3 ± 0.7 Å (70)	4.0 Å

The residue for which the measurement was acquired, the placement of the isotopic label, the experimentally determined distance (Å) or angle (°), ensemble statistics at the {001}, {010}, {100}-T1, {100}-T2, and {101} surfaces from round 2 biased predictions with a constraint weight *w* of 1 kcal/mol (values are average distances or angles from the 100 structures with lowest $E_{constraint}$, errors are standard deviation, values in parenthesis indicate number of structures from the 100 structures with lowest $E_{constraint}$ that met the measurement within experimental error), and corresponding prediction from the representative model (Rep. Model).

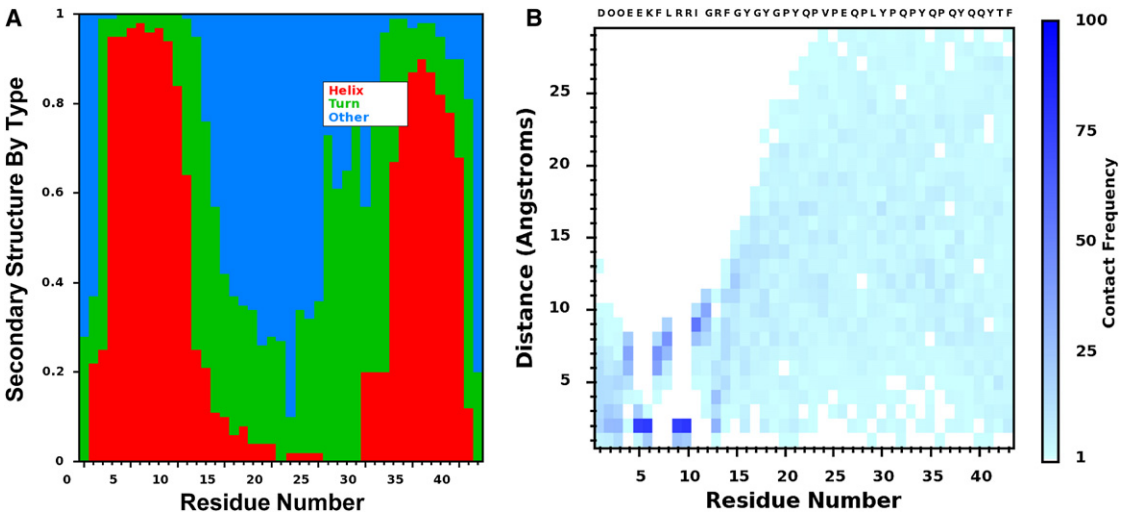

Figure 4. Ensemble Structure of Statherin Adsorbed to HAp {001} Surface Determined from the Final Phase of Biased Structure Determination
(A) Distributions of three basic secondary structure motifs and (B) pairwise residue-surface distances for the 100 predicted structures with lowest $E_{constraint}$. The structural designations "Helix" and "Turn" were assigned using the RosettaSurface hydrogen-bond function. The structural designation "Other" indicates that hydrogen bonding was either long range or absent at that residue.

ensemble statistics presented in Figure 4a: the 100 structures with smallest $E_{constraint}$ have helical N- and C-terminal segments, and the proline-rich middle segment has greater dispersion. Because the 100 structures with smallest $E_{constraint}$ adopt diverse middle-segment structures, the global RMSD of each structure is large. The method used here, of comparing all overlapping fragments (see Figure 6 legend), is useful for showing residue-specific structural variation while adsorbed to the surface.

Table 4 shows the effect of experimental constraints and the constraint weight on the calculated values for selected statherin-HAp distances. Glu5 is closer to the surface in round 2, compared with round 1, owing to the inclusion of that experimental constraint. When the constraint weight is increased for the final phase of structure determination, Glu5 moves closer to the surface and within the experimental error of the measurement. Round 2 calculations locate Arg13 and Phe14 closer to the

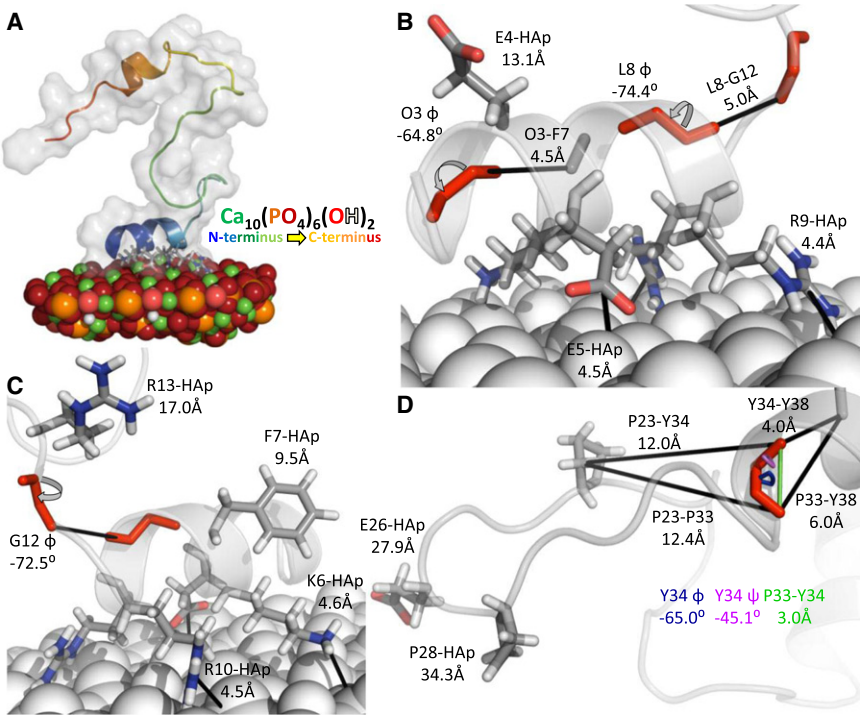

Figure 5. The Molecular Structure of Statherin Adsorbed to HAp {001} Surface
(A) Representative structure from the final phase of round 2 biased predictions. Opacity represents statherin's molecular shape, cartoons represent secondary structure, and sticks are shown for amino acids that are known from experiment to interact with HAp. Predicted distance and angle measurements at constrained atoms for the N- (B) and (C) and C-terminal domains (D). For comparison with ssNMR measurements see Table 3. In (D), three measurements are color coded for clarity. The one-letter amino-acid code for phosphoserine (Sep) is *O*.

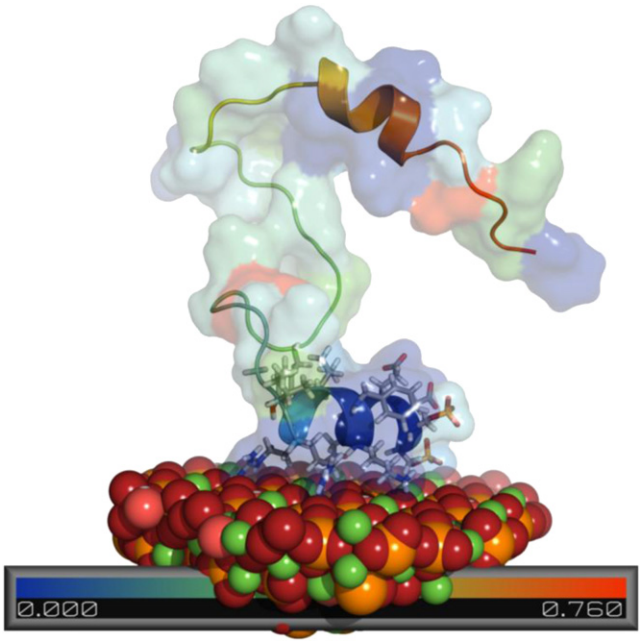

Figure 6. Local Root-Mean-Square Deviation of a Representative Structure Relative to the 100 Structures with Lowest Constraint Energy

To determine variation among structures, we superimposed the C^{α} atoms of every overlapping 3-mer in the representative structure and each of the 100 structures with lowest $E_{constraint}$, and calculated the root-mean-square deviation of the central C^{α} atom in the 3-mer. The color bar is in angstroms and corresponds to the opaque profile defining statherin's shape.

surface, compared with Round 1 calculations, even though those experimental constraints place Arg13 and Phe14 off the surface; this is because Glu5, Arg9, and Arg10 are experimentally biased to bind HAp in round 2, resulting in a more tightly bound N-terminal segment in general. When the constraint weight is increased for the final phase of structure determination, a statherin conformation is adopted allowing Glu5, Arg13, and Phe14 measurements all to be in their experimentally determined range.

DISCUSSION

The advent of X-ray crystallography and solution NMR has advanced our understanding of biomolecules. The structures determined from these methods could be considered the foundation of molecular biology. From the rational design of disease therapeutics to understanding natural strategies in evolved phenotypes, these structures are instrumental for understanding the molecules of life. Membrane biophysicists commonly acknowledge the disparity of solved membrane protein structures relative to those available for globular proteins, and there are >180 unique membrane protein structures solved by X-ray crystallography (White, 2009). For those studying the biophysics of biomineral-associated proteins, there are no solved structures, and this is perhaps the biggest detriment to the field of protein biomineralization. In this study, we have explored the potential to determine the structure of a biomineral-adsorbed

Table 4. Representative Predicted Statherin-HAp Distances from the Three Phases of Structure Determination

	E_5,C^{δ} (4.2 ± 0.3)	R13,C^{ζ} (>7)	F14,C^{ζ} (>7)
Round 1	8.2 ± 2.5	8.1 ± 3.8	13.4 ± 4.1
Round 2	5.0 ± 0.5	7.6 ± 2.5	12.4 ± 2.8
Final	4.5 ± 0.2	11.4 ± 3.8	13.8 ± 3.4

Experimentally determined value are in parentheses, computed values are average distances from the 100 structures with lowest $E_{constraint}$ with errors as standard deviation, and all values are Å.

protein by iterating between solid-state NMR measurements and protein structure prediction. The process converged rapidly on structural motifs in the N and C-terminal region, and while the models still retain significant dispersion in the middle section of the protein.

Protein X-ray crystal and solution-NMR structures are reported with varying certainty and a varying number of measurements per residue. The certainty of the models determined by these methods can depend on molecular weight, experimental conditions, or experimental capabilities at the time of determination. Recent solid-state NMR approaches have been used to determine amyloid structures by using three to five measurements per residue (Petkova et al., 2002), membrane channels by using one orientational measurement per residue (Hu et al., 2007), and membrane associated proteins by using zero to two measurements per residue (Sun and Weliky, 2009). In this study, we have used 22 measurements on a 43-residue protein, or just over one measurement for every two residues. These measurements have been obtained by overcoming the challenges of synthesizing and labeling many charged and phosphorylated residues as commonly found in biomineralization peptides. The constraints help yield a set of models which are remarkably consistent in the N- and C-terminal regions, but have considerable dispersion in the central portion of the protein. It is often thought that biomineralization proteins are typically flexible when associated with the mineral (Hunter et al., 2010). With the underdetermination of the data, it is still unclear whether the protein is dynamic and flexible (and thus represented by the family of models) or simply underdetermined.

Proteins can affect the formation of biominerals by directly binding a crystal face (Boskey et al., 1993; Elhadj et al., 2006; Kensuke and Yoshiki, 2001; Shiraga et al., 1992; Sollner et al., 2003). Because the structure of interacting biomolecules influences function and mechanism, determining the structure of biomineral-associated proteins is necessary for understanding how proteins influence biomineralization. Little high-resolution structural data is available for protein solid-surface complexes. In the absence of this data, structural models based entirely on computation are difficult to validate. Similarly, it is difficult to build structural models based on minimal high-resolution experimental data without the use of computation. The combined method developed here uses a structure prediction algorithm to search conformation space deemed relevant by high-resolution experimental data. These techniques appear to be an excellent complement, resulting in structural models with increased resolution, certainty, and scope.

Particularly encouraging is that fact that most ssNMR measurements used for biased predictions (Table 3) had been

predicted using RosettaSurface in the absence of experimental bias. The a priori predictions included the preferential adsorption of Glu5 relative to Glu4, the adsorption of basic residues Lys6, Arg9, and Arg10, the helical structure of the binding domain, and the lack of adsorption in the C-terminal domain. This predictive ability accelerates the convergence between experiment and computation and suggests that iterative structure determination using the combined approach could begin with structure prediction rather than ssNMR experiment. Beginning with structure prediction could reduce the amount of ssNMR data required for structure determination by informing the initial placement of isotopic labels.

Previously, unbiased RosettaSurface simulations on full-length statherin predicted no interaction between HAp and Phe14 (Masica and Gray, 2009). Previous ssNMR experiments on a statherin fragment truncated at residue 15 showed binding of Phe14 (Gibson et al., 2006). We concluded that Phe14 may locate at the charged HAp surface in the truncated experimental construct because of its proximity to the charged carboxy terminus (Masica and Gray, 2009). Results from this work show that Phe14 does not interact with HAp in full-length statherin.

Arg13 was predicted to adsorb to HAp in all RosettaSurface predictions before round 2 biased RosettaSurface.NMR predictions (Makrodimitris et al., 2007; Masica and Gray, 2009); however, ssNMR measurements show Arg13 to be away from the surface. The absence of binding via Arg13 may be surprising considering its proximity to the charged HAp surface in statherin's helical HAp-binding domain. Previously, we predicted that statherin's helical binding domain oriented its four basic residues to complement the geometry of periodic phosphate clusters at the HAp {001} surface (Makrodimitris et al., 2007). In a subsequent study we predicted that Arg13 adsorbed more tightly at the HAp {001} surface, compared with the {010} and {100} surfaces, owing to the geometry of periodic phosphate clusters at those surfaces (Masica and Gray, 2009). Given that HAp in tooth enamel grows along the c-axis (i.e., by deposition in the {001} plane) (Simmer and Fincham, 1995) and that statherin inhibits HAp growth (Schwartz et al., 1992), it is reasonable that one of the biologically relevant HAp crystal faces for statherin adsorption is the {001} surface. In stable HAp crystallites, like those used in the ssNMR experiments, the increased surface area of the {010} and {100} surfaces is predicted to reduce binding via Arg13.

With the advent of this combined computational-ssNMR approach, analysis can now be made at a sufficient resolution to begin to understand residue- and atom-specific contributions to the process of biomineralization and hard tissue formation. Analysis of high-resolution adsorbed-state protein structures will allow us to answer questions concerning phase-boundary biophysics. The results presented here indicate that statherin has a stable, folded HAp-binding domain, and the models suggest possibly few long-range contacts outside the binding domain. The methods combined here show significantly greater congruency at three of five tested HAp crystal surfaces, suggesting some specificity. A test of the success and general applicability of the combined method will require investigation on other protein biomineral systems. Although conformational sampling demands currently limit the method to proteins under ~50 amino acids, the approach should be generally applicable to proteins that can be chemically synthesized or biosynthetically labeled and surfaces that incorporate atomic species with ½-spin nuclei.

EXPERIMENTAL PROCEDURES

Protein Synthesis and Adsorption

Amino acids with [99%-^{13}C] and [98%-^{15}N] isotopes were purchased from Cambridge Isotope Laboratories and were incorporated using solid phase peptide synthesis based on fluorenylmethyloxycarbonyl-chemistry. The protein was purified and analyzed as described before (Goobes et al., 2006b). The protein was prepared in a phosphate buffer (100 mM NaCl, 40 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, pH 7.4) to maintain constant pH conditions at temperatures below the freezing point of water. A total of 16 mg of protein were adsorbed by mixing 30 mg of the protein in phosphate buffer with 100 mg of hydroxyapatite in the same phosphate buffer for 4 hr. After centrifugation the wet pellet containing the complex was washed three times with the phosphate buffer. The last wash contained only traces of protein in the buffer. The mineral-protein complex was then flash-frozen with liquid nitrogen before insertion into a precooled NMR probe to increase cross-polarization efficiency while avoiding water crystallization or salt depletion. From previous studies, no substantial changes in the secondary structure of the protein's N terminus were observed on freezing the protein-mineral complex (Long et al., 2001). In another work, statherin showed no structural changes at temperatures above ambient (Elgavish et al., 1984).

NMR Experiments

^{13}C-^{13}C reference measurements and ^{13}C-^{19}F rotational-echo double-resonance (REDOR) measurements were carried out on a home-built wide-bore 500 MHz spectrometer using a Varian 4 mm ^{1}H-^{19}F-^{13}C triple-tuned probe. The ^{13}C-^{13}C measurement was carried out at a spinning rate of 5 kHz and the ^{13}C-^{19}F REDOR measurement (S_0 and S) was carried out at a spinning rate of 8 kHz to achieve better signal/noise ratios. Carbon cross-polarization was achieved using a ramped field between 46 kHz and 23 kHz on the protons and a field of 35 kHz on the carbon with a contact time of 2 ms. Carbon 180° pulses and fluorine 180° pulses at respective fields of 45 kHz and 62 kHz were used. Continuous wave decoupling at a field of 85 kHz was used during the whole experiment. ^{13}C-^{15}N REDOR measurements were carried out on a Bruker DSX300 spectrometer using a 4 mm ^{1}H-^{13}C-^{15}N triple-tuned probe. Carbon cross polarization was achieved using a ramped field between 55 kHz and 27 kHz on the protons and a field of 40 kHz on the carbon with a contact time of 2 ms. Carbon 180° pulses and nitrogen 180° pulses at respective fields of 56 kHz and 30 kHz were used. Proton decoupling at a field of 75 kHz utilizing the two pulse phase modulation sequence was used to decouple the protons through the experiments. CN-REDOR experiments were carried out at a spinning rate of 5000 Hz and CF-REDOR experiments were done at a spinning rate of 8000 Hz. All experiments were carried out at a temperature of –50°C. ^{13}C-^{31}P REDOR measurements were carried out on a home-built wide-bore 500 MHz spectrometer using a Varian 4 mm ^{1}H-^{31}P-^{13}C triple-tuned probe operating at a spinning speed of 6 kHz. Carbon cross-polarization was achieved using a ramped field between 46 kHz and 23 kHz on the protons and a field of 35 kHz on the carbon with a contact time of 1 ms. Continuous wave decoupling at a field of 100 kHz was used during the whole experiment. Carbon 180° pulses and phosphorus 180° pulses at respective fields of 70 kHz and 60 kHz were used. Experiments were carried out at a temperature of –32°C. Cooled nitrogen gas flow was used for sample spinning at 5000 Hz and to achieve low temperatures at the sample. Dried and cooled air was used to cool the sample spun at 8000 Hz on the 500 MHz spectrometer. Reference and dephasing REDOR experiments were all carried out using the XY8 phase cycling to correct for inhomogenous radio frequency fields and other pulse imperfections.

Spin Dynamics Simulations

Calculations were carried out using the SIMPSON (Bak et al., 2000) spin dynamics program. For spin-pair simulations all chemical shift anisotropy and dipolar interaction parameters were taken into account. Chemical shift anisotropy parameters were computed from analysis of sideband patterns in

cross-polarization magnetic-angle spinning experiments and were confirmed against values from the literature (Ye et al., 1993). Known directions of the chemical-shift anisotropy (CSA) tensor components relative to the bonding of the carbonyl carbon in the backbone frame (Tycko et al., 1996) were used to generate all possible carbonyl-carbonyl relative CSA orientations (Euler angles) based on rotations around the torsion angles (ϕ,ψ) for the $\phi < 0$ section in the Ramachandran plot. For spin-triad simulations all interaction parameters of the ^{13}C spins and all dipolar interaction parameters of the heteronuclear dipolar interactions were taken into account. The geometry of the heteronuclear dipolar interactions was computed relative to [1-^{13}C]Y34 CSA principle axes system by placing the heteroatom at variable distances from the two carbon atoms and rotating the dipolar vectors around the C-C vector to cover all conformational space. $^{13}C_2$-X (X = F or N) simulations used all premeasured ^{13}C-^{13}C CSA and dipolar parameters.

Data Analysis

^{13}C-^{13}C simulations with (ϕ(r_{C-C}), ψ) as variables were used to create a grid of expected dephasing curves for allowed backbone conformations in the Ramachandran plot ($\phi < 0$). Calculated $\chi^2(\phi,\psi)$ function was used to fit experimental data. $^{13}C_2$-X (X = F or N) simulations with the two heteronuclear distance vectors and their orientation in the PAS of the [^{13}C]-Y34 CSA as variables were used to create an array of expected REDOR curves. These simulations used all measured ^{13}C-^{13}C CSA and dipolar parameters. Simulations of REDOR curves between protein side chain ^{13}C and ^{15}N spins were performed as described in Goobes et al. (2007) and references cited therein. Simulations of known distributions of ^{15}N and ^{31}P spins in diammonium hydrogen phosphate, for example, show that accurate heteronuclear dipolar couplings to the two nearest neighbor ^{31}P spins can be extracted from a truncated ^{31}P dephasing system assuming a ^{31}P-^{31}P dipolar coupling of 600 Hz. Following this strategy, simulations of the REDOR data using both an isolated ^{13}C-^{31}P spin pair model, and a ^{13}C-^{31}P-^{31}P triad.

The Algorithm

The algorithm developed here is based on RosettaSurface (Masica and Gray, 2009). Briefly, each execution of the RosettaSurface algorithm folds a protein from a fully extended conformation in solution using a united-atom model. Then, a high-resolution (all-atom) representation of the peptide is refined in solution and adsorbed to a biomineral surface in a random orientation. The fold and orientation of the protein are refined on the surface resulting in one candidate adsorbed-state structure. High-resolution refinement includes backbone, side-chain, and rigid-body optimization using Newtonian minimization.

Development of RosettaSurface.NMR included two modifications: First, the full-atom energy (E) used for decoy discrimination is a linear combination of attractive and repulsive Lennard-Jones interactions (E_{att} and E_{rep}), solvation (E_{sol}), hydrogen bonding (E_{hb}), electrostatics (E_{coul}), and a term to enforce ssNMR constraints ($E_{constraint}$):

$$E = E_{att} + E_{rep} + E_{sol} + E_{coul} + E_{hbond} + wE_{constraint} \quad (1)$$

where:

$$E_{constraint} = \sum_{i=1}^{n} E_{jconstraint} \quad (2)$$

where:

$$E_{jconstraint} = \begin{cases} 0, \text{ if } x_{jmeasured} - \sigma < x_{jpredicted} \\ \text{and} \\ x_{jmeasured} + \sigma > x_{jpredicted} \\ \text{else} \\ \left(\dfrac{x_{jmeasured} - x_{jpredicted}}{\sigma_i} - 1\right)^2 \end{cases} . \quad (3)$$

In Equation 3, $x_{jmeasured}$ and $x_{jpredicted}$ are the i^{th} perimentally determined angle or distance and predicted angle or distance, respectively; σ_i is the experimental error of the i^{th} measurement; n is the number of biasing constraints used during structure prediction (Equation 2). For residues experimentally determined to not bind the surface (Glu4, Phe7, Arg13, Phe14, Glu26, and Pro28), $E_{jconstraint} = 0$ for predictions within one standard deviation of the experimental measurement—for predictions closer to the surface than one standard deviation, $E_{jconstraint}$ increases according to an exponential, fit to the experimental data. During biased structure prediction, the constraint weight w (Equation 1) modulates the strength of the bias toward experimental measurements; the units of the weight are kcal/mol. Because the RosettaSurface energy function used for structure generation in this study includes a constraint energy term, a true binding energy is not calculated.

A second modification to RosettaSurface is that immediately after the formation of the adsorbed-state complex, the protein is moved randomly in an attempt to satisfy all experimentally determined protein-surface intermolecular contacts. Random moves consist of a Gaussian distributed translation of mean 0.1 Å in each Cartesian direction and a Gaussian distributed random rotation of mean ~17° around each Cartesian axis. The algorithm makes 100 random moves of this type. Each move that reduces the magnitude of $E_{constraint}$ (for n measured protein-surface intermolecular constraints) is accepted, otherwise the complex is returned to its previous state (i.e., zero temperature Monte Carlo search). During this intermolecular constraint optimization, the surface is represented as a plane to avoid calculating every interatomic distance. After 100 attempts the protein-surface system proceeds to adsorbed-state refinement and surface atoms are returned to their explicit representations.

For all predictions, 10^5 candidate HAp-adsorbed statherin structures were generated using RosettaSurface.NMR. Each candidate structure is assessed using the constraint energy $E_{constraint}$ (for all n ssNMR measurements). Structures with the smallest constraint energy are used for further analysis and model representation.

Statherin Model

For a starting structure, we built an extended molecular structure of human salivary statherin using PyMOL (DeLano, 2002). The atomic parameter set is the same as in Masica and Gray (2009).

Hydroxyapatite Model

We built monoclinic hydroxyapatite crystals using CrystalMaker (Palmer, 2001). In total, five HAp surfaces were used in this study (see Figure 2): {001}, {010}, two differentially terminated {100} faces ({100}-T1 and {100}-T2), and {101}. The {100}-T2 surface terminates with a higher density of calcium ions than the {100}-T1 surface. All other HAp surfaces were cut to expose approximately neutral (mixed charged) surfaces. The atomic parameter set is the same as in Masica and Gray (2009).

The RosettaSurface software is currently licensed for free to users at academic and nonprofit institutions at http://graylab.jhu.edu/.

ACKNOWLEDGMENTS

D.L.M. and J.J.G. were supported by The Arnold and Mabel Beckman Foundation's Young Investigator Award, and the National Science Foundation's CAREER Award (0846324). J.T.A., M.N., and G.P.D. were supported by the National Institutes of Health (R01 GM074511, R01 DE12554).

Received: July 19, 2010
Revised: September 10, 2010
Accepted: September 17, 2010
Published: December 7, 2010

REFERENCES

Astala, R., and Stott, M.J. (2008). First-principles study of hydroxyapatite surfaces and water adsorption. Phys. Rev. B *78*, 75427–75438.

Bak, M., Rasmussen, J.T., and Nielsen, N.C. (2000). SIMPSON: a general simulation program for solid-state NMR spectroscopy. J. Magn. Reson. *147*, 296–330.

Boskey, A.L., Maresca, M., Ullrich, W., Doty, S.B., Butler, W.T., and Prince, C.W. (1993). Osteopontin-hydroxyapatite interactions in vitro: inhibition of hydroxyapatite formation and growth in a gelatin-gel. Bone Miner. *22*, 147–159.

Bowers, P.M., Strauss, C.E.M., and Baker, D. (2000). De novo protein structure determination using sparse NMR data. J. Biomol. NMR *18*, 311–318.

Cavalli, A., Salvatella, X., Dobson, C.M., and Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. Proc. Natl. Acad. Sci. USA *104*, 9615–9620.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Dorozhkin, S.V., and Epple, M. (2002). Biological and medical significance of calcium phosphates. Angew. Chem. Int. Ed. Engl. *41*, 3130–3146.

Dove, P.M., De Yoreo, J.J., and Weiner, S. (2003). Biomineralization, *Volume 54* (Washington, DC: Mineralogical Society of America, Geochemical Society).

Dussol, B., Geider, S., Lilova, A., Leonetti, F., Dupuy, P., Daudon, M., Berland, Y., Dagorn, J.C., and Verdier, J.M. (1995). Analysis of the soluble organic matrix of five morphologically different kidney stones. Urol. Res. *23*, 45–51.

Elgavish, G.A., Hay, D.I., and Schlesinger, D.H. (1984). 1H and 31P nuclear magnetic resonance studies of human salivary statherin. Int. J. Pept. Protein Res. *23*, 230–234.

Elhadj, S., De Yoreo, J.J., Hoyer, J.R., and Dove, P.M. (2006). Role of molecular charge and hydrophilicity in regulating the kinetics of crystal growth. Proc. Natl. Acad. Sci. USA *103*, 19237–19242.

Gibson, J.M., Raghunathan, V., Popham, J.M., Stayton, P.S., and Drobny, G.P. (2005). A REDOR NMR study of a phosphorylated statherin fragment bound to hydroxyapatite crystals. J. Am. Chem. Soc. *127*, 9350–9351.

Gibson, J.M., Popham, J.M., Raghunathan, V., Stayton, P.S., and Drobny, G.P. (2006). A solid-state NMR study of the dynamics and interactions of phenylalanine rings in a statherin fragment bound to hydroxyapatite crystals. J. Am. Chem. Soc. *128*, 5364–5370.

Gong, H., Shen, Y., and Rose, G.D. (2007). Building native protein conformation from NMR backbone chemical shifts using Monte Carlo fragment assembly. Protein Sci. *16*, 1515–1521.

Goobes, G., Goobes, R., Schueler-Furman, O., Baker, D., Stayton, P.S., and Drobny, G.P. (2006a). Folding of the C-terminal bacterial binding domain in statherin upon adsorption onto hydroxyapatite crystals. Proc. Natl. Acad. Sci. USA *103*, 16083–16088.

Goobes, R., Goobes, G., Campbell, C.T., and Stayton, P.S. (2006b). Thermodynamics of statherin adsorption onto hydroxyapatite. Biochemistry *45*, 5576–5586.

Goobes, G., Stayton, P.S., and Drobny, G.P. (2007). Solid state NMR studies of molecular recognition at protein-mineral interfaces. Prog. Nucl. Magn. Reson. Spectrosc. *50*, 71–85.

Gray, J.J. (2004). The interaction of proteins with solid surfaces. Curr. Opin. Struct. Biol. *14*, 110–115.

Hu, J., Asbury, T., Achuthan, S., Li, C., Bertram, R., Quine, J.R., Fu, R., and Cross, T.A. (2007). Backbone structure of the amantadine-blocked transmembrane domain M2 proton channel from influenza A virus. Biophys. J. *92*, 4335–4343.

Hunter, G.K., O' Young, J., Grohe, B., Karttunen, M., and Goldberg, H.A. (2010). The flexible polyelectrolyte hypothesis of protein-biomineral interaction. Langmuir, in press. 10.1021/la100401r.

Long, J.R., Shaw, W.J., Stayton, P.S., and Drobny, G.P. (2001). Structure and dynamics of hydrated statherin on hydroxyapatite as determined by solid-state NMR. Biochemistry *40*, 15451–15455.

Makrodimitris, K., Masica, D.L., Kim, E.T., and Gray, J.J. (2007). Structure prediction of protein- solid surface interactions reveals a molecular recognition motif of statherin for hydroxyapatite. J. Am. Chem. Soc. *129*, 13713–13722.

Mann, S. (2001). Biomineralization. Principles and Concepts in Bioinorganic Materials Chemistry (Oxford, England: Oxford University Press).

Masica, D.L., and Gray, J.J. (2009). Solution-and adsorbed-state structural ensembles predicted for the statherin-hydroxyapatite system. Biophys. J. *96*, 3082–3091.

Meiler, J., and Baker, D. (2003). Rapid protein fold determination using unassigned NMR data. Proc. Natl. Acad. Sci. USA *100*, 15404–15409.

Kensuke, N., and Yoshiki, C. (2001). Control of crystal nucleation and growth of calcium carbonate by synthetic substrates. Chem. Mater. *13*, 3245–3259.

Ndao, M., Ash, J.T., Breen, N.F., Goobes, G., Stayton, P.S., and Drobny, G.P. (2009). A (13)C{(31) P} REDOR NMR investigation of the role of glutamic acid residues in statherin-hydroxyapatite recognition. Langmuir *25*, 12136–12143.

Ndao, M., Stayton, P.S., and Drobny, G.P. (2010). The role of basic amino acids in the molecular recognition of hydroxyapatite by statherin using solid-state NMR. Surf. Sci. *604*, L39–L42.

Palmer, D. (2001). CrystalMaker. A computer program for plotting crystal structures. (http://www.crystalmaker.com).

Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F., and Tycko, R. (2002). A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. Proc. Natl. Acad. Sci. USA *99*, 16742–16747.

Tycko, R., Weliky, D.P., and Berger, A.E. (1996). Investigation of molecular structure in solids by two-dimensional NMR exchange spectroscopy with magic angle spinning. J. Chem. Phys. *105*, 7915–7930.

Rohl, C.A., and Baker, D. (2002). De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. J. Am. Chem. Soc. *124*, 2723–2729.

Ryall, R.L. (1996). Glycosaminoglycans, proteins, and stone formation: adult themes and child's play. Pediatr. Nephrol. *10*, 656–666.

Schwartz, S.S., Hay, D.I., and Schluckebier, S.K. (1992). Inhibition of calcium phosphate precipitation by human salivary statherin: structure-activity relationships. Calcif. Tissue Int. *50*, 511–517.

Shaw, W.J., Ferris, K., Tarasevich, B., and Larson, J.L. (2008). The structure and orientation of the C-terminus of LRAP. Biophys. J. *94*, 3247–3257.

Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J.M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K.K., and Lemak, A. (2008). Consistent blind protein structure generation from NMR chemical shift data. Proc. Natl. Acad. Sci. USA *105*, 4685–4690.

Shiraga, H., Min, W., VanDusen, W.J., Clayman, M.D., Miner, D., Terrell, C.H., Sherbotie, J.R., Foreman, J.W., Przysiecki, C., and Neilson, E.G. (1992). Inhibition of calcium oxalate crystal growth in vitro by uropontin: another member of the aspartic acid-rich protein superfamily. Proc. Natl. Acad. Sci. USA *89*, 426–430.

Sigel, A., Sigel, H., and Sigel, R.K.O. (2008). Biomineralization. From Nature to Application, *Volume 4* (West Sussex, England: Wiley).

Simmer, J.P., and Fincham, A.G. (1995). Molecular mechanisms of dental enamel formation. Crit. Rev. Oral Biol. Med. *6*, 84–108.

Sollner, C., Burghammer, M., Busch-Nentwich, E., Berger, J., Schwarz, H., Riekel, C., and Nicolson, T. (2003). Control of crystal size and lattice formation by Starmaker in otolith biomineralization. Science *302*, 282–286.

Stayton, P.S., Drobny, G.P., Shaw, W.J., Long, J.R., and Gilbert, M. (2003). Molecular recognition at the protein-hydroxyapatite interface. Crit. Rev. Oral Biol. Med. *14*, 370–376.

Sun, Y., and Weliky, D.P. (2009). 13C-13C correlation spectroscopy of membrane-associated influenza virus fusion peptide strongly supports a helix-turn-helix motif and two turn conformations. J. Am. Chem. Soc. *131*, 13228–13229.

White, S.H. (2009). Biophysical dissection of membrane proteins. Nature *459*, 344–346.

Ye, C., Fu, R., Hu, J., Hou, L., and Ding, S. (1993). Carbon-13 chemical shift anisotropies of solid amino acids. Magn. Reson. Chem. *31*, 699–704.